

What X-ray source counts can tell about the large scale matter distribution

A. M. Soltan and M. J. Chodorowski

Nicolaus Copernicus Astronomical Center, Bartycka 18, 00-716 Warsaw, Poland
e-mail: soltan@camk.edu.pl, michal@camk.edu.pl

Received / Accepted

ABSTRACT

Context. Sources generating most of the X-ray background (XRB) are dispersed over a wide range of redshifts. Thus, statistical characteristics of the source distribution carry the information on the matter distribution on very large scales.

Aims. We test the possibility to detect the variation of the X-ray source number counts over the celestial sphere.

Methods. A large number of *Chandra* pointings spread over both galactic hemispheres is investigated. A search for all the point-like sources in the soft band of 0.5–2 keV is performed, and statistical assessment of the population of sources below the detection threshold is carried out. A homogeneous sample of the number counts at fluxes above $\sim 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ for more than 300 ACIS fields was constructed. The counts correlations between overlapping fields were used to assess the accuracy of the computational methods used in the analysis.

Results. It is shown that the source number counts vary between fields at the level only slightly larger than the fluctuation amplitude expected for the random (Poissonian) distribution. Nevertheless, small asymmetry between galactic hemispheres is present. The average number of sources in the northern hemisphere is larger than in the southern at the 2.75σ level. Also the autocorrelation function of the source density in both hemispheres are substantially different.

Conclusions. Possible explanations for the observed anisotropies are considered. If the effect is unrelated to the observational selection, a large scale inhomogeneities in the distribution of X-ray sources are required. Correlations of the source number counts observed in the southern hemisphere could be generated by a coherent structure extending over $\sim 1200 \text{ Mpc}$.

Key words. X-rays: diffuse background – X-rays: general – large-scale structure of Universe

1. Introduction

The largest structures identified in galaxy redshift surveys are located at distances comparable to the maximum distance at which structures can be effectively distinguished. In the CfA redshift survey beyond the Great Wall not much structure is at all recognizable (Geller & Huchra 1989). In the SDSS¹ the large galaxy filaments are discernible up to roughly 500 Mpc, and the most prominent feature is the Sloan Great Wall (Gott et al. 2005), found at a distance of $\sim 300 \text{ Mpc}$.

Larger structures potentially could be isolated in the quasar part of the SDSS. Recently Clowes et al. (2012) report a group of 73 quasars that span redshifts 1.17–1.37. A total length of this twisted filament exceeds 1200 Mpc. Albeit Park et al. (2012) have shown that largest structures found in the galaxy distribution are consistent with the canonical ΛCDM model, it is plausible that the quasar alignment could be discordant with this model.

It appears that the question of the large structures could be addressed using the X-ray surveys. Most of extragalactic X-ray sources are associated with various types of Active Galactic Nuclei (AGN). Because of a strong cosmic evolution and a wide luminosity function, the AGN observed in the selected flux range are distributed within a very wide range of redshifts. Equivalently, the observed source flux is only weakly correlated with the source distance. Roughly 90 percent of the X-ray background in the 0.5–2 keV energy band is generated

by sources brighter than $3 \cdot 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ (e.g. Moretti et al. 2003). A majority of these sources are located at cosmological distances: 80 percent in the redshift range 0.3–2.2 (Soltan 2008). The comoving radial distance between redshifts 0.3 and 2.2 exceeds 4100 Mpc^2 . Thus, sources detected in a moderately deep *Chandra* observation are tracers of the matter distribution at Gpc scales.

In the deepest *Chandra* fields of 2 and 4 Ms exposure several hundred sources are detected. This allows to investigate the source distribution along the line of sight, provided the redshifts of objects are known. However, to investigate large scale matter distribution, a transverse dimension of the survey should be of a comparable size as the radial one. In the present paper we explore the possibility to analyze the giant structures in the observable Universe using a large number of the *Chandra* medium depth observations scattered over the whole sky. Although a single pointing represents a pencil-like cut through space, a set of several hundred pointings to some extent act as a uniform all-sky survey. The main objective was to select from the *Chandra* Data Archive all the pointings suitable for source detection, and then to construct an efficient method to count the sources in all the fields in a homogeneous way. Number of sources in each field was used to estimate the amplitude of the source counts, $N(S)$, at a fixed flux range around $10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. A power law parametrization of the $N(S)$ relationship was applied. Statistical characteristics of the $N(S)$ amplitude variations are examined.

Send offprint requests to: A. M. Soltan

¹ <http://www.sdss.org/>

² Assuming “737” ΛCDM cosmology, i.e. $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

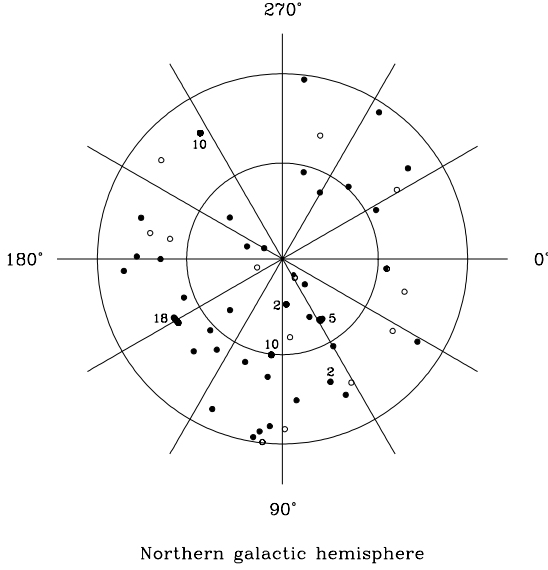


Fig. 1. The distribution of 80 pointings in the northern galactic hemisphere. Two circles represent galactic latitudes $b = 30^\circ$ and $b = 60^\circ$. Each observation covers a circle of 10 arcmin in diameter and the symbol sizes are not to scale. Labels close to some points in the map denote the actual number of pointings in the area. All the pointings are detached. Full dots indicate pointings with high statistical weights (see Sect. 4).

Results, albeit preliminary, demonstrate strong and weak points of the method. The volume of the archive *Chandra* observations is still too small for a comprehensive study. Nevertheless, the presently available data reveal the potential ability of X-ray source counts in the investigation of the large scale matter distribution.

The organization of the paper is as follows. In the next section, selection and preparation of the observational data extracted from the *Chandra* archives is described. In Sect. 3, procedures to estimate the amplitude of the number source counts, $N(S)$, are presented. Two independent methods to determine the counts in two separate flux levels are used. Results of these calculations, as well as statistical and (possible) systematic errors are discussed in Sects. 4 and 5. The main conclusion are summarized in Sect. 6.

2. Selection of observational material

The *Chandra* Data Archive has been searched for all the ACIS-I observations at high galactic latitudes with the exposure time above 30 ks. Originally, a selection limit of $|b| > 30^\circ$ has been applied for both hemispheres. Then, because of a small number of adequate pointings, a search for more observations in a belt of $25^\circ < b < 30^\circ$ in the southern galactic hemisphere has been performed. Eventually, within this extension only 4 pointings with $27^\circ < b < 30^\circ$ satisfied all the selection criteria (see below) and were included into the final set of data.

All the “interesting” observations were handled through the standard data processing pipelines at the *Chandra* X-ray Center³. Each observation was carefully inspected in terms of

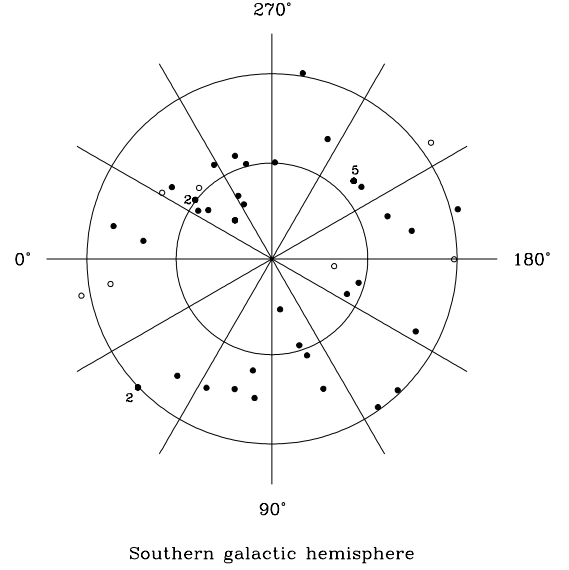


Fig. 2. Same as Fig 1 for 41 pointings in the southern galactic hemisphere.

usefulness for the present investigation. Pointings centered at galactic clusters, nearby galaxies (e.g. Magellanic Clouds or NGC galaxies) have been discarded. Also pointings with dominating extended source have been removed. If the extended source occupied a relatively small fraction of the field of view, a section with the source was cut-off, and the remaining data unaffected by the extended emission have been used. Fields with extremely strong point sources have been treated in a similar way, i.e. only areas affected by linear smearing generated during a readout have been removed. All the exposures were examined for the presence of the background flares and only “good time intervals” (GTI) were used. Because of a severe deterioration of the image quality with increasing off-axis angle, search for sources was restricted to the circular area of 5 arcmin radius centered on chips 0 – 3.

More than 300 individual pointings have been qualified for the analysis. Their spatial distribution is highly nonrandom. Concentrations in various angular scales, as well as the strong asymmetry between galactic hemispheres are present. Since a large number of pointings covered overlapping areas of the sky, the data have been divided into two samples. The detached pointings, which cover separate areas, constitute a principal sample used in the investigation. The second auxiliary sample consisting of overlapping pointings was used to assess uncertainties of our statistical analysis.

The sky distribution of the non-overlapping pointings is presented in the polar equal area projection in Figs. 1 and 2 for the north and south galactic hemispheres, NGH and SGH, respectively. Symbols in figures strongly exceed in size the actual extent of the individual observation. Numerals close to some dots denote numbers of pointings concentrated in the marked areas. In the NGH 106 pointings have been selected for further processing, while in the SGH - 49. Full dots represent pointings characterized by high statistical weights (80 in NGH and 41 in SGH); open circles - observations that suffer from the largest statisti-

³ See <http://asc.harvard.edu/ciao>.

cal uncertainties. A detailed discussion of statistical properties of the selected material is presented in Sect. 4.

The data in a “canonical” soft energy band of 0.5 – 2 keV have been used. Because of various reasons, both the exposure time and sensitivity are highly variable function of position within the field of view. To reduce effects of sensitivity variations, low sensitivity areas have been delineated using the a *Chandra* effective exposure map⁴. For each observation, the maximum amplitude of the exposure map was determined. Pixels with the exposure map below 75 % of this maximum value have not been used. Although this constraint reduces the original area of 5 arcmin circle by roughly 10 %, it secures for our purpose an acceptable uniformity of the exposure (Softan 2011).

3. ‘Bright’ and ‘faint’ source counts

Two independent methods have been applied to estimate the number of discrete point-like sources in a single observation. The first one was based on a standard procedure of finding prominent concentrations of counts within a detection cell. Sources isolated in this way are labeled here as ‘bright’. A second method, described in Softan (2011), estimates number of the ‘faint’ sources. It calibrates deviations of the photon distribution from the Poissonian one using the nearest neighbor statistics. It is not able to isolate individual sources, but provides statistical information on the population of sources generating small number of photons.

To count ‘bright’ sources in a single observation, a circular cell was slid over the entire investigated area. The cell size was set to enclose 85 % of counts generated by a point source. Radius of the detection cell varied as a function of the off axis angle according to the shape of the *Chandra* point spread function (PSF). The PSF data given by Jerius et al. (2000) have been used. The search for ‘bright’ sources has been performed in the same manner for all the pointings. The iterative procedure was applied. Using the sliding box technique, the strongest concentration of photons was localized and it was recognized as the brightest source in the field of view. Then, the area surrounding this source was extracted from the data and the search for the second brightest source was performed in the same way. The procedure was repeated for the consecutive sources until the number of counts selected as the next source dropped below the threshold adopted individually for each observation. The threshold counts were selected at the level sufficiently high to prevent random fluctuation of counts to be approved as a real source. It was assumed in this step that the background counts were distributed randomly over the field of view. The search for sources has been terminated when the probability of random accumulation of counts exceed 10^{-5} .

Our data sets span a wide range of exposure times, from 30 ks through more than 170 ks. Because of that the pointings differ in the source detection threshold and in the number of detected sources. To use the observed number of sources as the estimator of the source counts per unit solid angle, which is independent of the exposure time, one needs to adopt an analytic model for the counts and to relate the observed number of sources in the individual field to the parameter(s) of the model.

A power law is the obvious choice for the source number counts per unit solid angle (hereafter per 1 sq. deg):

$$\frac{dN(S)}{dS} = N_o S^a, \quad (1)$$

with two a priori free parameters, normalization amplitude N_o and the slope a . However, the limited number of sources as well as a narrow dynamic range of fluxes populated by sources prohibits us from attempts to measure the slope a . Also the data on the source counts available in the literature for a couple of deep fields do not provide information on the potential field to field slope variations. In those few cases where the dynamic range within a single field allows to track changes of the slope, it appears that the slope is constant over the whole range covered by the present investigation (e.g. Georgakakis et al. 2008). Therefore, in the present investigation only the normalization is fitted for each field separately, while a single, universal value of the slope is assumed.

A several step procedure is applied to find the best estimate of N_o using the observed number sources. Because the each enhancements of counts within the detection cell is considered a ‘source’, one should carefully evaluate a contribution of the background counts and apply the adequate statistical correction. Let $n_s(k)$ is the number of detected ‘sources’ containing k counts, where $k = k_s + k_b$ is a sum of the genuine X-ray photons emitted by a source k_s , and the background counts k_b . N_o is related to $n_s(k)$ in the following way:

$$n_s(k) = A \sum_{k_b=0}^k \left[P_{\text{bkg}}(k_b) \int dS \frac{dN(S)}{dS} P(k_s | S) \right], \quad (2)$$

where A is the solid angle of the observation (in sq. deg), $P_{\text{bkg}}(k_b)$ denotes the probability that k_b background counts would be found in the detection cell, and $P(k_s | S)$ is the probability that a source of flux S generates k_s photons (see Eq. 4). The probability of finding k_b counts within the detection cell is given by Poissonian distribution:

$$P_{\text{bkg}}(k_b) = \frac{e^{-\lambda} \lambda^{k_b}}{k_b!}, \quad (3)$$

where λ is the expected number of random counts within the detection cell.

The sum over the background counts k_b in Eq. 2, generally includes the $k_s = 0$ component. However, the minimum value of k was set to warrant $P_{\text{bkg}}(k_b = k) < 10^{-5}$, and the last component of the sum is effectively negligible. It is convenient to use the instrumental count in each observation as a flux unit. The flux s in ACIS counts is related to S in physical units by a conversion factor η : $s = S/\eta$, with η in $\text{erg s}^{-1} \text{cm}^{-2} \text{count}^{-1}$ is obtained from the exposure map. The probability $P(k_s | S)$ now takes the form:

$$P(k_s | s) = \frac{e^{-s} s^{k_s}}{k_s!}. \quad (4)$$

Substituting into Eq. 2 the source number counts in new units

$$\frac{dn(s)}{ds} = n_o s^a, \quad (5)$$

where

$$n_o = A N_o \eta^{a+1}, \quad (6)$$

⁴ An exposure map is generated in a standard data processing. It is a position dependent product of the effective area of the mirror/detector combination and the effective exposure time [in cm^2s].

and assuming the constant slope a over a wide flux range, we finally get:

$$n_s(k) = n_o \sum_{k_b=0}^{k-1} \left[P_{\text{bkg}}(k_b) \frac{\Gamma(k_s + a + 1)}{\Gamma(k_s + 1)} \right]. \quad (7)$$

Equation 7 summed over k gives the expected total number of ‘bright’ sources. To evaluate n_o , the number of actually detected ‘bright’ sources has been substituted in the left-hand side of Eq. 7. The amplitude N_o was obtained using Eq. 6. The statistical properties of the N_o distribution in our sample are discussed in the next section in conjunction with the analogous distribution obtained for ‘faint’ sources.

The population of sources below the formal source detection threshold, labeled as ‘faint’, is investigated using a statistical approach. Here only the basics of the method are presented. All the details and potential applications were discussed in a separate paper (Softan 2011).

After the removal of all the ‘bright’ sources, the remaining counts are a mix of the non-X-ray events, truly diffuse X-ray background and counts generated by a population of faint sources. Only this last constituent is responsible for fluctuations of the count distribution characteristic for point sources. One can express the total number of events, n_t , within the investigated area in the following form:

$$n_t = n_1 + n_2 + \dots + n_{k_{\text{max}}}, \quad (8)$$

where n_1 represents counts distributed randomly, n_2 - counts produced by sources contributing exactly 2 counts each (‘2-photon sources’), n_3 - counts by ‘3-photon sources’ and so on up to $n_{k_{\text{max}}}$ - counts due to the brightest sources left in the field (i.e. the brightest among the ‘faint’ ones). Counts distributed randomly constitute a composite collection of events that includes also the weakest discrete sources contributing single photons. Photons coming from sources producing $2 \leq k \leq k_{\text{max}}$ counts create local enhancement of a unique shape determined by the telescope PSF. Statistical characteristics of these variations are efficiently quantified using the nearest neighbor statistics (NNST) of the count distribution. The probability that randomly chosen event has no neighbors within a distance r :

$$P(r) = p_1 P(r|1) + p_2 P(r|2) + \dots + p_{k_{\text{max}}} P(r|k_{\text{max}}), \quad (9)$$

where p_1 is a probability that this event is not related to any source generating 2 or more counts, p_k for $k = 2, \dots, k_{\text{max}}$ is the probability that this event is produced by the k -photon source, and $P(r|k)$ is the conditional probability that there are no other counts within r provided the selected event belongs to the k -photon source. Under the reasonable assumptions, for $k \geq 2$:

$$P(r|k) = P(r|1) \cdot \mathcal{P}(r|k), \quad (10)$$

where $\mathcal{P}(r|k)$ is the probability that within r there are no other counts from the same source. This quantity is uniquely defined by the PSF. Taking $n_k = k \cdot n_s(k)$ and $p_k = n_k/n_t$ and combining Eqs. 9, 10, 8, and 5 we get:

$$\frac{n_o}{n_t} P(r|1) \sum_{k=2}^{k_{\text{max}}} \frac{\Gamma(k + a + 1)}{\Gamma(k)} [1 - \mathcal{P}(r|k)] = P(r|1) - P(r). \quad (11)$$

Equation 11 allows us to determine n_o provided the relevant probability distributions are known. One can estimate the probability $P(r)$ for each observation by measuring the distance to the nearest neighbor for each photon in the field. Similarly, $P(r|1)$ is

given by the distribution of distances to the nearest photon from the randomly distributed points. The PSF related probability distribution $\mathcal{P}(r|k)$ was investigated in detail by Softan (2011). It has been determined as a function of the off axis angle using several *Chandra* observations of bright point sources.

The best estimate of n_o was determined in a following way. All the probability distributions have been calculated for several (usually more than 20) distances r , which uniformly covered a domain of distributions $P(r|1)$ and $P(r)$. Then, the cumulative probability distributions in Eq. 11 were replaced by the corresponding probability densities, e.g. $\Delta P(r) = P(r) - P(r + \Delta r)$ for the consecutive distances r . In all the calculations $\Delta r = 0.246$ arcsec ($\equiv 1/2$ of the instrumental pixel) has been used. Thus, a set of independent equations covering all the observed separations has been constructed, and the best fit value of n_o was determined using the least square method. Finally, the count amplitude N_o was obtained using Eq. 6 as for the ‘bright’ sources.

4. Counts correlations

The amplitudes of the number source counts determined for ‘bright’ and ‘faint’ sources (hereafter denoted $N_o[b]$ and $N_o[f]$, respectively) define the number source counts over adjoining, but different flux levels. Therefore, the median flux of ‘bright’ and ‘faint’ sources in each observation are not widely separated and $N_o[b]$ and $N_o[f]$ should be similar. Nevertheless, one can expect systematic differences between both estimates, if the count slope adapted in the calculations, a , did not match the slope actually observed in the data. The differential slope between 10^{-16} and 10^{-14} erg s $^{-1}$ cm $^{-2}$ in the deep *Chandra* fields is well fitted with the slope of -1.58 (Georgakakis et al. 2008; Softan 2011). Using $a = -1.58$ in the present calculations, the average amplitude of the ‘faint’ source counts is larger by $\sim 5\%$ than that for the ‘bright’ sources. Both amplitudes are practically identical for $a = 1.60$. To preserve internal consistency of the investigation, the latter slope has been used in the subsequent calculations. Although this is steeper by 0.02 than the best fit by Georgakakis et al. (2008), it stays within their 1σ limits. One should note that the present investigation is not suitable to assess the count slope either within the individual field or the average over the sky. This is because most systematic errors involved in the estimates of the number of ‘bright’ sources are independent from the errors affecting the ‘faint’ sources assessment. The source counts amplitudes, $N_o[b]$ and $N_o[f]$ are functionally dependent with the slope a . Therefore, a unique slope is required to balance both amplitudes. In fact, modification of the slope by just 0.02 demonstrates that the systematics do not play a significant role in the analysis. One should stress also that all the conclusions of the paper are not affected by a selection of the particular value of the slope.

Apart from the systematic errors, the $N_o[b]$ and $N_o[f]$ amplitudes are subject to large statistical uncertainties that generate substantial scatter. The average number of ‘bright’ sources per field $\bar{n}_s = 23.6$. Thus, the Poissonian scatter alone introduces $\sim 21\%$ rms fluctuations to the present estimate of $N_o[b]$. In reality, still larger variations of $N_o[b]$ are expected due to source clustering. The N_o estimates for the ‘faint’ sources are afflicted by even larger errors, because the estimate of the count amplitudes is derived indirectly. Two factors contribute most to the final uncertainties. First, the amplitude estimate is proportional to the difference between two probability distributions (see Eq. 11), and both distributions are subject to substantial statistical errors. Second, the ‘faint’ sources contribution to the total events is small, typically 2 – 6%, while all the remaining counts are dis-

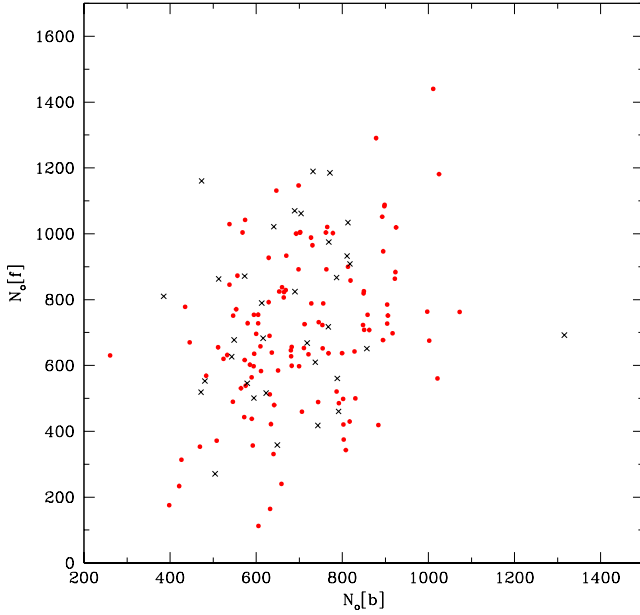


Fig. 3. Amplitudes N_o of the source number counts for flux S in $10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ (see Eq. 1) for the ‘bright’ and ‘faint’ sources in the sample of 155 pointings; 121 pointings with the *combined* rms uncertainty of N_o smaller than 134 per sq. deg are shown with dots, 34 pointings with larger rms - with crosses.

tributed randomly. Therefore the net signal, represented by the right-hand side in Eq. 11, is also very modest in comparison to the noise that is dominated by the random scatter of the $P(r|1)$ distribution.

In Fig. 3 the ‘bright’ and ‘faint’ source amplitudes N_o as defined in Eq. 1, where S is in $10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$, are plotted for the full sample of the non-overlapping 155 pointings. The rms scatter of $N_o[b]$ is in fact comparable to that expected from the Poissonian fluctuations of the ‘bright’ source numbers, while the $N_o[f]$ variance is substantially larger. The rms for ‘faint’ sources amounts to 34 % of the average amplitude of $N_o[f]$. Despite the scatter, ‘bright’ and ‘faint’ amplitudes are strongly correlated. The correlation coefficient for 155 data points $\rho = 0.2968$. The non-directional probability that the data are drawn from the uncorrelated general population amounts to 0.00018 (or 0.000089 for the directional case). To minimize statistical noise, the uncertainties of the individual N_o estimates are carefully examined (see below). Due to a wide range of exposure times in the processed observations, these uncertainties strongly differ between each other. In order to check the influence of the poor quality estimates on the subsequent statistical analysis, pointings with the largest uncertainties have been identified. In Fig. 3 these pointings are shown with crosses.

Although the $N_o[f]$ estimates are strongly affected by the statistical noise at the level of the photon distribution within the each pointing, significant correlation between $N_o[f]$ and $N_o[b]$ in the whole sample favors the calculations of a single amplitude N_o derived simultaneously from the ‘bright’ and ‘faint’ sources. In order to combine the $N_o[b]$ and $N_o[f]$ estimates into a single ‘best’ estimate of the source number counts amplitude for each pointing, an adequate error estimate and data weighting for each measurement are essential. The errors for the ‘bright’ source counts have been calculated using the Poissonian noise, since

this effect is dominating. In the first step it was assumed that the underlying population of sources is the same for all the pointings. Next, the expected number of the observed ‘bright’ sources has been determined for each pointing separately taking into account the exposure time. Finally, the Poissonian distribution around the expected number of sources was used to compute the corresponding dispersion $\sigma_o(N_o[b])$ individually for each pointing.

The $N_o[f]$ uncertainties result from a complex combination of the Poisson-like fluctuations of the number of faint sources, contribution of background counts, and a stochastic character of the photon distribution at angular scales comparable to the width of the PSF. In effect, the variance of $N_o[f]$ cannot be easily derived from theoretical considerations, but instead it is estimated directly from the observed distribution. No clear correlation of the variance with the exposure time (or the total number of events) has been found. It was evident that the cosmic variance contributed marginally to the observed $N_o[f]$ scatter. Therefore, equal uncertainties have been assigned to all the observations. As an error of the individual $N_o[f]$ estimate, $\sigma_o(N_o[f])$, the rms scatter in the whole sample had been assumed: $\sigma(N_o[f]) = 242$ per sq. deg.

Our final best estimate of N_o is a weighted mean of $N_o[b]$ and $N_o[f]$ with weights inversely proportional to the squares of the rms uncertainties of both estimates. The resulting uncertainties of N_o , $\sigma_o(N_o)$, are spread between 101 and 149 per sq. deg. To reduce an influence of the low statistical quality data for further analysis, observations with the highest uncertainties have been removed from the sample. A threshold of $\sigma_o(N_o) = 134$ per sq. deg has been set as a compromise between the acceptable uncertainties and the size of the sample. The resultant sample of 121 observations is shown with dots in Fig. 3, and the remaining 34 pointings with $\sigma_o(N_o) > 134$ per sq. deg – with crosses. A removal of 34 observations that suffer from the largest uncertainties substantially increases the correlation between $N_o[b]$ and $N_o[f]$. For 121 data points $\rho = 0.3589$. The corresponding non-directional probability for the uncorrelated general population is $5 \cdot 10^{-5}$. Unless otherwise stated, the discussion below refers to the sample of 121 observations

The average number counts amplitude $\bar{N}_o = 705$ per sq. deg. with the rms scatter $\sigma(N_o) = 145$ per sq. deg. The flux limits that define ‘bright’ and ‘faint’ sources depend on the exposure time, but the flux separating both source classes is always close to $10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. Therefore the present \bar{N}_o estimate characterizes the source counts just around this flux level.

To compare the present result with the source number counts based on several deep *Chandra* fields we adopt the count parametrization by Georgakakis et al. (2008). Their data converted to the units used here give $dN/dS = 574$ per sq. deg at $S = 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$, i.e. our figure is 23 % higher. Although this difference partially could be attributed to various systematic effects inherent to the present calculations, one should note that our amplitude is the average of a large number of pointings distributed over the whole sky, while the *Chandra* Deep Fields cover few selected areas. Applying the present method to the selection of 54 observations in the CDFS gave $N_o = 601 \pm 88$, per sq. deg, what exceeds the Georgakakis et al. (2008) result merely by 5 %. The quoted uncertainty represents the rms scatter among the 54 observations. Although all these data are localized at CDFS, the fields cover slightly different areas. Therefore variations of N_o stand not just for the statistical noise generated by our procedures, but include also actual fluctuations in the number of sources

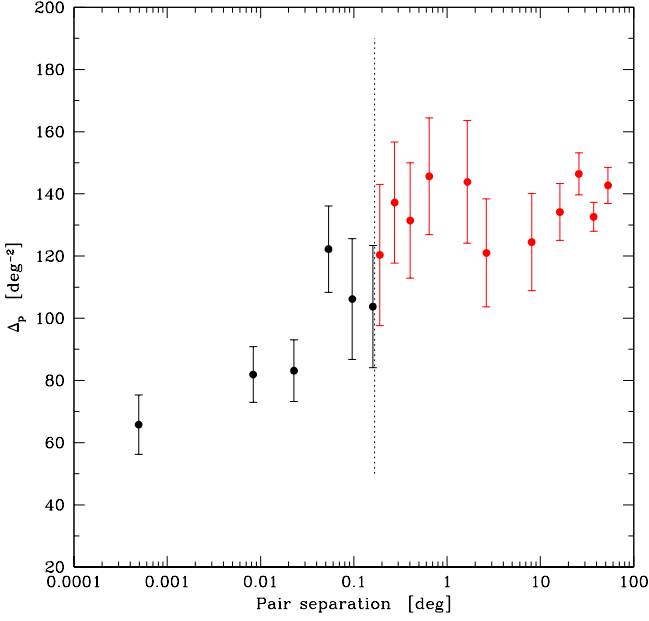


Fig. 4. The distribution of the ‘pair rms’, i.e. rms difference between the counts amplitudes in pairs of observations as a function of the pair separation (see text for detailed explanation). The dotted line separates the overlapping pointings from the detached ones. The lowest separation bin contains pairs closer than 0.1 arcmin.

To assess the uncertainties genuine to the method, and clean from the contribution generated by the cosmic variance, all 264 observations (including overlapping) have been used in the following way. For all the possible pointing pairs a difference between amplitudes N_o was obtained: $\Delta_{ij} = N_o(i) - N_o(j)$, for $i, j = 1, \dots, 264$, and $i \neq j$. The pairs were divided in several bins according to the pair angular separation. Then, for the each separation bin the rms of Δ_{ij} has been calculated:

$$\sigma_p^2 = \frac{1}{N_{ij}} \frac{1}{2} \sum_{i \neq j} \Delta_{ij}^2, \quad (12)$$

where N_{ij} denotes the number of pointing pairs in the given separation bin, and ‘2’ in the denominator is introduced to normalize σ_p to the rms of N_o . For perfectly overlapping pointings, the ‘pair rms’ σ_p represents the scatter intrinsic to the method, while for widely separated fields, σ_p is a quadratic sum of the intrinsic scatter and the cosmic variance. If the observations partially overlap, intermediate amplitudes of σ_p are expected. Potentially, for close, but non-overlapping fields a correlation signal generated by the large scale structures could also be present.

In Fig. 4 the distribution of σ_p for several bins is shown. The abscissa positions give the average pair separation in the bin. The lowest separation bin accommodates pairs for which the field centers are separated by less than 0.1 arcmin. The vertical dotted line separates the overlapping fields from the detached ones. One sigma error bars were calculated using the bootstrap Monte Carlo method. A large number (10000) of simulated sets of observations were generated. The original sky coordinates have been used and a random amplitude N_o was drawn from the the full assembly of 121 N_o estimates. Then, using Eq. 12 the pair rms σ_p have been calculated for each simulated data set.

The rms scatter of the σ_p ’s was taken as the uncertainty of the actually observed signal.

The subsample of overlapping observations is dominated by few groups centered on *Chandra* Deep Fields. Some widely separated bins which happen to include the separation between these fields contain pairs predominantly drawn from the selected areas. To eliminate this bias, the Δ_p for the separations larger than 10 arcmin have been determined using the sample of the non-overlapping observations alone.

A substantial decline of the pair rms for overlapping pointings clearly indicates diminishing contribution of the cosmic variance. It is legitimate to assume that at the smallest separations the amplitude of $\sigma_p = 65.8 \pm 9.6$ per sq. deg represents the scatter of N_o due to imperfections of the applied procedures and is independent from the varying cosmic signal. Since the observed rms of the N_o distribution in the sample of 121 non-overlapping observations amounts to $\sigma_{N_o} = 144.8$ per sq. deg, the intrinsic variance of the counts amplitude $\sigma_{CV} = 129.0^{+4.5}_{-5.3}$ per sq. deg, or $(18.3^{+0.6}_{-0.8})\%$ of the average N_o amplitude. Apparently the Poissonian fluctuations provide a dominant contribution to the observed scatter.

5. Counts in the north and south galactic hemispheres

No obvious variations of the pair rms is present for the non-overlapping pointings averaged over the whole data. It indicates that any potential strong fluctuations of the counts amplitude have angular scales that typically do not exceed a size of the individual pointing. Nevertheless, some differences in statistical properties of the N_o distribution between both galactic hemispheres seem to be present.

Unfortunately, the observations useful for the present analysis are not distributed evenly in both hemispheres. Of 121 pointings in the basic sample, 80 is found in the northern hemisphere, and 41 in the southern. The average amplitude and the rms scatter of the number counts in the northern and southern hemispheres are $\overline{N_o(N)} = 731.6 \pm 135.3$ and $\overline{N_o(S)} = 654.7 \pm 150.5$. Assuming that the uncertainty of the average scales as one over square root of the number of observations, we have $\sigma[\overline{N_o(N)}] = 15.1$ and $\sigma[\overline{N_o(S)}] = 23.5$. Then, the uncertainty of the difference between both averages equals to 27.9 and the signal-to-noise of the difference reaches 2.75. The K-S two sample test applied to the both N_o distributions indicates that the northern and southern data sets differ at the confidence level of 0.95. Thus, global properties of both hemispheres reveal some differences, but at rather low significance.

Further details of the N_o distribution are revealed using the autocorrelation functions (ACF). These functions have been calculated in both hemispheres separately over a wide range of angular distances, and the results are shown in Fig. 5. Data for the northern hemisphere (shown with crosses) are consistent with no correlation signal for the non-overlapping fields, i.e. at angular scales above 10 arcmin. The ACF uncertainties have been assessed using the Monte Carlo method in the same way as for the σ_p . The ACF in the observed part of the SGH (shown with dots in Fig. 5) exhibits distinctly different characteristics. The correlations apparent here indicate substantial large angular scale variations of the count amplitude. The positive ACF amplitude seems to extend up to $\sim 30^\circ$ and it is conspicuously inconsistent with the ACF in the northern cap (the negative ACF signal at larger separations arises in a natural way because the size of whole investigated SGH region is just few times larger than the

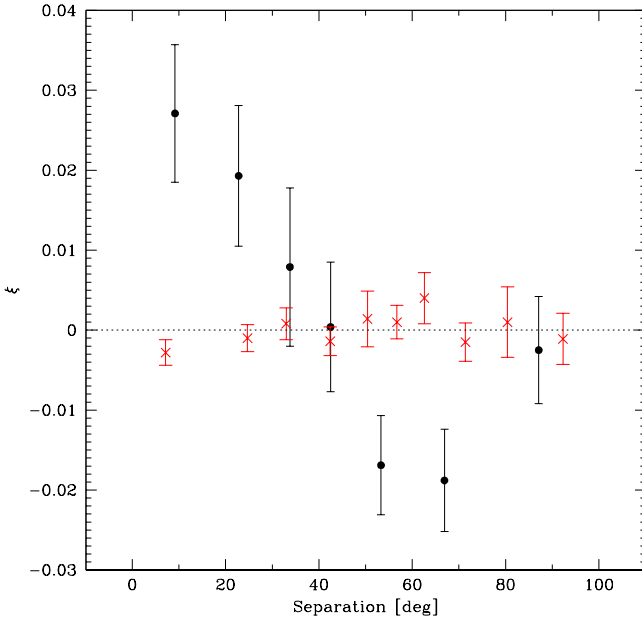


Fig. 5. The autocorrelation function for the N_o distribution in the northern (crosses) and southern (dots) hemisphere. The error bars show 1σ uncertainties.

apparent correlation length). The 1σ uncertainties in the SGH data are larger than in the NGH due to smaller number of observations and larger rms scatter of the N_o amplitudes in the former hemisphere.

6. Discussion

Variations of the source number counts have been analyzed using the *Chandra* observations distributed in both galactic hemispheres. The counts amplitude was obtained using two independent and complementary methods. The dominating contribution in the observed scatter comes from the Poissonian noise induced by the relatively small average number of sources per field. This effect is seriously confusing a search for the potential correlations in the N_o distribution. Scarce observational data, particularly in the south galactic hemisphere, prevent us from drawing conclusions at high confidence level. The pointings are distributed highly non homogeneously in the celestial sphere in the large as well as in small angular scales. Due to these deficiencies, the picture emerging from the present investigation is to some extent ambiguous. Fluctuations of the counts amplitude averaged over the whole sample of 121 observations do not exhibit any irregularities which could be symptomatic for the existence of the large angular scale structure of the XRB. However, statistical properties of the data split into the northern and southern hemispheres reveal noticeable differences between the source distributions. In the northern hemisphere the ACF amplitude does not exceed a few thousandths and is within errors consistent with zero. This result is in agreement with our investigation (Soltan et al. 1996) based on the *ROSAT* All-Sky Survey (RASS). We analyzed a section of the NGH above $b = 40^\circ$ with $70^\circ < l < 250^\circ$ in the energy band $0.73 - 1.56$ keV. The RASS ACF amplitude in that region drops to $\sim 2 \cdot 10^{-3}$ at separations $\theta \approx 1^\circ$ and below $5 \cdot 10^{-4}$ at $\theta > 5^\circ$.

A distinctly different picture emerges from the present investigation in the SGH. Here at separations below $\sim 30^\circ$ the ACF amplitude rises to ~ 0.02 and differs from zero with significance above 3σ . Close inspection of the N_o distribution in both hemispheres indicates that the ACF amplitude in the SGH is generated by a group of pointings with lower than average N_o amplitudes in the longitude quadrant $240^\circ < l < 330^\circ$. It is clear that the ACF signal determined in the SGH is discordant with the ACF in the NGH using the present data as well as with the ACF of the RASS determined again from a section of the NGH. The effect undoubtedly deserves further study. One should note, that the apparent differences between hemispheres cannot be explained by inhomogeneities of the soft X-ray absorption by cold gas within the Galaxy. To generate a 12% difference in the number of sources in both hemispheres, hydrogen column densities should differ by almost 10^{21} cm^{-2} , what is excluded by the observational data.⁵

Although all the pointings were carefully examined, the correlations in the N_o distribution could be a result of a conspiracy of the instrumental effects or observational selection. Nevertheless, a genuine cosmic signal should not be ruled out. The average counts amplitude of 14 pointings in the SGH between longitudes 240° and 330° is lower by 26% than the average amplitude of all the remaining 107 pointings. Such decline of the source counts could be realized by a single void or by a radial alignment of several voids. In the case of the single void, one can estimate its line of sight extent using the model of the source redshift distributions by Soltan (2008). The minimum size of 1200 Mpc in comoving coordinates is obtained with the void center at redshift of ~ 0.8 . This size is comparable to the longest dimension of the quasar chain found by Clowes et al. (2012) in the SDSS. Transverse extents of the supposed void are rather weakly constrain due to poor sampling. Assuming that the subtended angle of the void corresponds to the correlation length of the ACF, the transverse size is of the same order as the radial one. Thus, the volume of the void would be much larger than the volume of the structure reported by Clowes et al. (2012).

A large number of the new pointings would improve statistical significance of the investigations. Unfortunately, one cannot expect a rapid inflow of the new *Chandra* observations. Therefore, a more promising way to proceed would be to search for the correlation between the existing *Chandra* observations and the data in other available X-ray surveys. In particular, the relationship between the counts amplitude determined here and the RASS SGH maps should be investigated.

Acknowledgements. I thank all the people who created the *Chandra* Interactive Analysis of Observations software for producing a user-friendly environment. This work has been partially supported by the Polish NCN grant 2011/01/B/ST9/06023.

References

- Clowes, R. G., Harris, K. A., Raghunathan, S. et al. 2012 arXiv:1211.6256
- Geller, M. J., & Huchra, J. P. 1989, *Science*, 246, 897
- Georgakakis, A., Nandra, K., Laird, E. S., Aird, J., & Trichas, M. 2008, *MNRAS*, 388, 1205
- Gott, J. R., Jurić, M., Schlegel, D., et al. 2005, *ApJ*, 624, 436
- Jerius, D., Donnelly, R. H., Yibbetts, M. S., et al. 2000, *Proc. SPIE*, 4012, 17
- Moretti, A., Campana, S., Lazzati, D., & Tagliaferri, G. 2003, *ApJ*, 588, 696
- Park, C., Choi, Y., Kim, J., et al. 2012, *ApJ*, 759, 7
- Soltan, A. M. 2008, *A&A*, 490, 1039
- Soltan, A. M. 2011, *A&A*, 532, A19
- Soltan, A. M., Hasinger, G., Egger, R., Snowden, S., & Trümper, J. 1996, *A&A*, 305, 17

⁵ See: <http://heasarc.gsfc.nasa.gov/cgi-bin/Tools/w3nh/w3nh.pl>